

Bayesian Nonparametric Learning of Cloth Models for Real-Time State Estimation

著者	Koganti Nishanth, Tamei Tomoya, Ikeda Kazushi, Shibata Tomohiro
journal or publication title	IEEE Transactions on Robotics
volume	33
number	4
page range	916-931
year	2017-05-17
その他のタイトル	Bayesian Nonparametric Learning of Cloth Models for Real-time State Estimation
URL	http://hdl.handle.net/10228/00007729

doi: info:doi/10.1109/TRO.2017.2691721

Bayesian Nonparametric Learning of Cloth Models for Real-time State Estimation

Nishanth Koganti^{1,2}, Tomoya Tamei¹, Kazushi Ikeda¹ and Tomohiro Shibata²

Abstract—Robotic solutions to clothing assistance can significantly improve the quality-of-life for the elderly and disabled. Real-time estimation of human-cloth relationship is crucial for efficient learning of motor-skills for robotic clothing assistance. The major challenge involved is cloth state estimation due to inherent non-rigidity and occlusion. In this study, we present a novel framework for real-time estimation of cloth state using a low-cost depth sensor making it suitable for a feasible social implementation. The framework relies on the hypothesis that clothing articles are constrained to a low-dimensional latent manifold during clothing tasks. We propose the use of Manifold Relevance Determination (MRD) to learn an offline cloth model which can be used to perform informed cloth state estimation in real-time. The cloth model is trained using observations from motion capture system and depth sensor. MRD provides a principled probabilistic framework for inferring the accurate motion-capture state when only the noisy depth sensor feature state is available in real-time. The experimental results demonstrate that our framework is capable of learning consistent task-specific latent features using few data samples and has the ability to generalize to unseen environmental settings. We further present several factors that affect the predictive performance of the learned cloth state model.

Index Terms—Personal Robots, Robotic Clothing Assistance, Visual Tracking, Learning and Adaptive Systems, Cloth State Estimation

I. INTRODUCTION

ASSISTIVE robots are playing an increasing role in improving the living standards and independence of the elderly and disabled population. Recent demographic trends including the tremendous increase of elderly population has caused a severe shortage of trained professionals for care-giving. The long term goal for the robotics community is the realization of care-giving robots that not only provide companionship but also physical assistance with everyday activities. The major requirements for such a care-giving robot includes safe and reliable human interaction and the ability to manipulate a wide array of household items. There has been significant research on rigid object manipulation, however, handling of non-rigid objects such as clothes remains a challenging task with active ongoing research.

Clothing assistance is a basic assistive activity in the daily life of the elderly and disabled. However, robotic clothing assistance is considered an open problem as it involves a

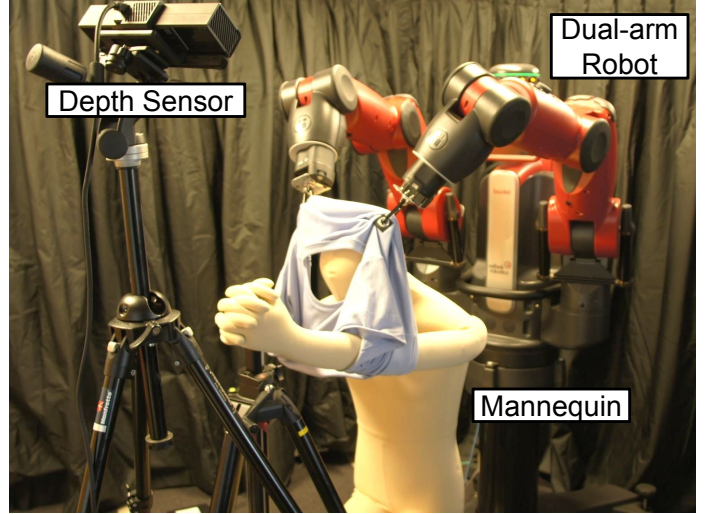


Fig. 1. Clothing assistance performed by dual-arm robot with cloth state estimation performed using depth sensor.

tightly coupled interaction among the human subject, non-rigid clothing articles and the robot. Real-time estimation of human-cloth relationship enables detection and adaptation to failure scenarios while performing clothing tasks, which is required for a practical implementation of a robotic framework. The challenge in this problem mostly lies with cloth state estimation due to their inherent non-rigidity, occlusion by the human subject and self-occlusion as seen in the experimental setup shown in Fig. 1.

Clothing articles inherently lie in a high dimensional configuration space and can undergo large deformations. The deformations occurring during a cloth folding task can be significantly different when compared to wearing the same clothing article. This makes general-purpose modeling and state estimation of clothing articles not only difficult but also impractical. However, clothing articles follow consistent deformations for a particular task thereby constrained to a low-dimensional manifold which is task-specific. A possible approach for reliable cloth state estimation is to constrain the search space within task-specific latent cloth models.

In this study, we propose the use of Bayesian nonparametric latent variable models to learn an offline cloth model to perform informed cloth state estimation in real-time as shown in Fig. 2. Our framework is applied for the real-time tracking of human-cloth relationship using a depth sensor. A depth sensor is a low-cost solution to capture 3D shape information without requiring any elaborate setup or calibration of the

¹Nishanth Koganti, Tamei Tomoya and Kazushi Ikeda are with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara, Japan nishanth-k, tomo-tam, kazushi@is.naist.jp

²Tomohiro Shibata is with the Graduate School of Life Sciences and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan tom@brain.kyutech.ac.jp

sensor. These features are crucial to develop a real-world implementation of clothing assistance wherein end-users and caregivers can easily setup the system. However, depth sensors provide noisy information and there is also the problem of cloth occlusion in our task setting. We tackle these problems by assuming that clothing articles are constrained to a low-dimensional latent manifold specific to clothing assistance tasks forming a task-specific cloth model. This model is learned using the non-linear dimensionality reduction technique Manifold Relevance Determination (MRD) [7] to handle the non-rigidity of clothing articles and learn the cloth latent features in a Bayesian manner avoiding the problem of over-fitting.

MRD is used to learn an offline low-dimensional latent manifold for data from simultaneous observation of clothing article using a motion capture system and a depth sensor. Both sensory systems have complimentary capabilities, when combined provide the most informative observation of clothing articles. The motion capture system can provide accurate location information of discrete markers in the environment, however, it is an expensive and complex system that requires precise calibration and can not be used in real-time. On the other hand, depth sensors are low-cost and calibration free, however, they provide noisy point cloud information of the whole environment. MRD provides a principled probabilistic framework for inferring the accurate motion capture state when only the noisy depth sensor state is available in real-time. In this study, we further investigate the effect of factors such as feature representations on the predictive performance of the trained cloth state model.

The rest of the paper is organized as follows. Section II, introduces some of the related works for robotic cloth handling. In Section III, we describe our proposed framework for learning cloth models. Section IV shows the experimental results. Finally we conclude in Section V with some future directions.

II. RELATED WORK

In the recent years, there has been a lot of attention on developing frameworks for robotic cloth manipulation. The studies tackle different aspects of the problem and are presented here in broadly three categories: *Cloth State Estimation*, *Motor-skills Learning*, *Robotic Clothing Assistance*. The first category has studies on reliable cloth state estimation. The second category includes motor-skills learning for non-rigid clothing articles. The third category presents studies that tackled robotic clothing assistance. This section further includes a literature survey on the use of latent manifold learning in the domain of robotics and computer vision.

Cloth State Estimation: One of the approaches to robotic cloth handling is to rely on efficient cloth state estimation along with static planning of the robot. Ramisa *et al.* [26], [27] proposed feature descriptors for detection and parts segmentation of clothing articles from RGB-D data. Willimon *et al.* [28], [29] used interactive perception with a robot to classify non-rigid objects in a cluttered environment. Kita *et al.* [30], [31] used multiple observations of a cloth to fit

a mesh-model through optimization used for informed cloth manipulation. These studies have several limitations in the context of robotic clothing assistance. The frameworks rely on the use of high-dimensional cloth state models and sometimes optimization based-techniques. Usually, planar assumption is taken for clothing articles to constrain state estimation. These assumptions are invalid for clothing assistance tasks and computationally efficient representations are required to ensure real-time state estimation. These studies also do not specifically handle the interaction between clothing articles and human-subjects required for clothing assistance.

Motor-skills Learning: Several studies have proposed motor-skills learning frameworks for cloth handling. Doumanoglou *et al.* [32] formulated a Partially Observable Markov Decision Process (POMDP) framework for cloth unfolding along with random forests for cloth classification. Huang *et al.* [33] generated trajectories through a warp function to bring clothes to a desired configuration. Lakshmanan *et al.* [34] used movement primitives to parametrize motion planning for performing a given sequence of folds. Miller *et al.* [24], [25] performed robust robotic cloth folding by generating motion trajectories for a 2D polygon approximation of the clothing article. The motor-skill learning frameworks presented here mainly handle tasks that require point-to-point planning based on one-shot cloth state estimation decisions. There are also studies that handle high-dynamics tasks, but the clothing articles do not undergo severe occlusions and the cloth state was represented by tracking specific positions of clothing articles such as the corners. However clothing assistance tasks are highly dynamical requiring efficient cloth state estimation to handle cloth occlusions and constraints due to coupling with the human.

Robotic Clothing Assistance: There are few studies that tackle the problem of clothing assistance. Klee *et al.* [35] proposed a clothing assistance framework to coordinate with a human to complete various clothing tasks. They emphasized on human motion tracking and performed tasks such as putting a cap on a human subject. Colome *et al.* [36] performed clothing of a mannequin with a scarf using reinforcement learning. They relied on an accurate inverse dynamics model for reliable motion planning. Gao *et al.* [37] tackled the problem through user-specific body constraints calibration to perform reliable motion planning for clothing assistance. Yamazaki *et al.* [38], [39] have proposed a framework for clothing of subjects with pants. In their framework, they relied on the use of optical flow and an offline database of image streams to detect the current state of clothing task. These studies presented here do not specifically handle reliable cloth state estimation and address other aspects of clothing assistance such as human-pose modeling and robot dynamics handling. The tasks in these studies also do not have much interaction between the clothing article and the human.

Manifold Learning in Robotics: Robotics applications require learning motor-skills with high-dimensional observations obtained using noisy sensors. Usually the inherent dimensionality of the task is much lower and is non-linearly related to the observation space. There have been several studies that use Gaussian Process (GP) based latent variable models (LVM) for

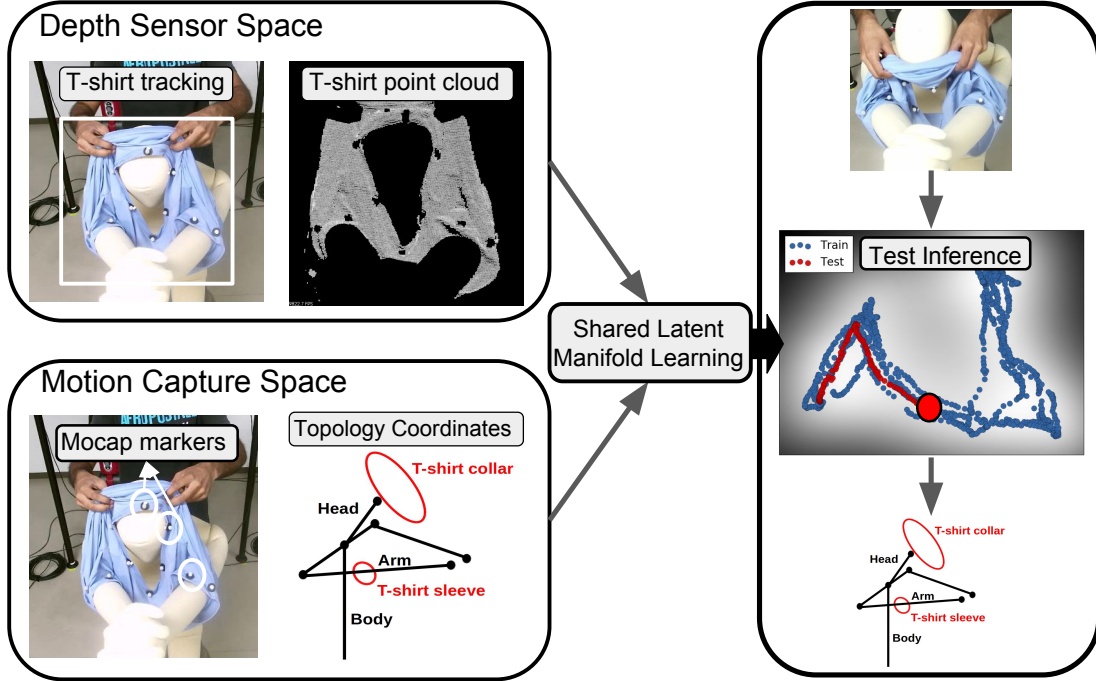


Fig. 2. Proposed framework for cloth state modeling using MRD. Observations are obtained from a depth sensor represented by cloth point cloud and from the motion capture system represented by topology coordinates. The learned latent manifold is used in real-time to infer human-cloth relationship information given noisy depth sensor input.

data-efficient learning. Shon *et al.* [40] formulated a shared GP latent variable model for multiple observation spaces and applied it to robotic imitation of human poses through a shared latent space. Ko *et al.* [42] formulated a generic framework for GP-Bayes filters in settings with incomplete ground truth data where a state sequence is generated using GPLVM. Dimensionality reduction has also been used in reinforcement learning frameworks for robotics where the policy learning is done using a low-dimensional latent space that captures task space constraints [41], [44]. Wang *et al.* [43] proposed a GPLVM based dynamics model that can infer the intention of a human subject given task demonstrations in a human-robot interaction setting. These studies rely on the use of GP based latent variable models in complex settings to perform data-efficient learning thereby validating its applicability to robotic clothing assistance.

In this study, we address the challenge where there is significant coupling between the clothing article and the human such as the clothing with a T-shirt involving severe cloth deformation and occlusion by the mannequin. Our study builds upon the clothing assistance framework proposed by Tamei *et al.* [3], wherein a dual-arm robot clothes a soft mannequin with a T-shirt. Topology coordinates were used as a low-dimensional state representation for efficient motor-skills learning. The proposed framework is an extension to our preliminary studies [1], [2] wherein we tackled the problem of reliable cloth state estimation using a depth sensor for clothing assistance tasks. In this study, we propose the offline fusion of a depth sensor and a motion capture system for reliable online tracking using a depth sensor. We further rely on the use of Gaussian process based non-linear dimensionality reduction

technique to handle the non-rigid dynamics of clothing articles. We also perform data-efficient learning through Bayesian nonparametric latent variable models so that the cloth state model can generalize to unseen environmental settings.

III. METHODS

In this section, we present our proposed method for the modeling of cloth state and its application to robotic clothing assistance tasks. First, the formulation of the cloth state model and its motivation is presented. We then describe the representations used for the depth sensor data (feature space) and motion capture data (pose space) in the cloth state model. Finally the different strategies implemented for the inference of pose state given test feature state is explained.

A. Non-linear Latent Manifold Learning

In this study, we assume that clothing articles undergo similar deformations during a specific type of task thereby constrained to a low-dimensional configuration space. We propose the use of Gaussian Process (GP) [4] based non-linear latent variable models [5], [7] to learn the underlying low-dimensional manifold from high dimensional cloth state observations. In this section, we provide the mathematical formulation of the models and discuss the applicability of these models for cloth state modeling.

Bayesian Gaussian Process Latent Variable Model (BG-PLVM) is a non-linear dimensionality reduction technique proposed by Titsias *et al.* [5]. It is derived from the generative model shown in Fig. 3a where the observations, $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$, $\mathbf{y}_n \in \mathbb{R}^D$, are assumed to be

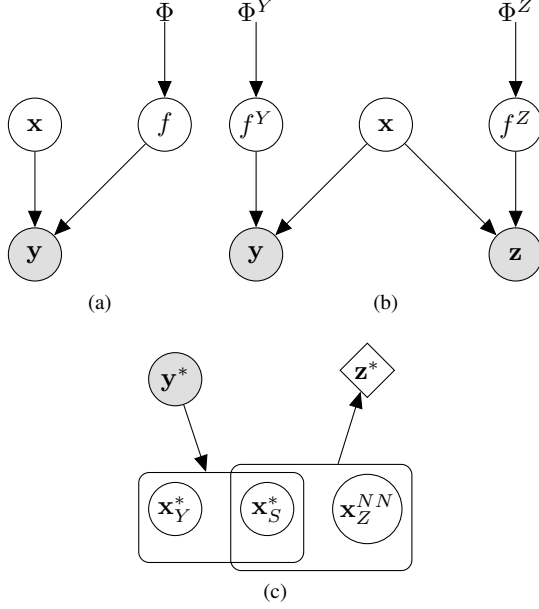


Fig. 3. Graphical models of latent variable models: a) Bayesian Gaussian Process Latent Variable Model (BGPLVM) [5], b) Manifold Relevance Determination (MRD) [7], c) Inference for MRD [7]

generated through a noisy process from latent variables $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, $\mathbf{x}_n \in \mathbb{R}^L$,

$$\mathbf{y}_n = f(\mathbf{x}_n) + \epsilon_n, \quad \epsilon_n \sim \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I}),$$

$$p(\mathbf{y}_n | \mathbf{x}_n, f, \beta) = \mathcal{N}(f(\mathbf{x}_n), \beta^{-1}\mathbf{I}) \quad (1)$$

where β denotes the inverse variance for the noise random variable ϵ and the conditional distribution for an observation sample can be derived as a Gaussian distribution. In this model, a prior on the mapping function f is placed using a Gaussian Process [4] $f(x) \sim \mathcal{GP}(\mathbf{0}, k(\mathbf{x}, \mathbf{x}'))$, where $k(\mathbf{x}, \mathbf{x}')$ is the covariance function. For performing automatic model selection of the latent space dimensionality, the Automatic Relevance Detection (ARD) kernel [4] can be used,

$$k_{\text{ARD}}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_{\text{ARD}}^2 \exp \left(-\frac{1}{2} \sum_{l=1}^L \alpha_l (x_{i,l} - x_{j,l})^2 \right) \quad (2)$$

The ARD weights $\{\alpha_l\}_{l=1}^L$ describe the relevance of each dimension and σ_{ARD} describes the scale of the GP mapping function. The relevance is usually determined using a heuristic threshold such that dimensions with weights below the threshold have insignificant contribution to reconstructing the observations [5], [7].

The objective is to infer the unknown latent variables \mathbf{X} and the model parameters $\Phi = \{\beta, \sigma_{\text{ARD}}^2, \{\alpha_l\}_{l=1}^L\}$ of the mapping function. The conditional likelihood is derived by assuming D independent GP mappings evaluated on the latent variables \mathbf{X} ,

$$p(\mathbf{Y}_{:,d} | \mathbf{X}, \Phi) = \mathcal{N}(\mathbf{Y}_{:,d} | \mathbf{0}, \mathbf{K}),$$

$$p(\mathbf{Y} | \mathbf{X}, \Phi) = \prod_{d=1}^D p(\mathbf{Y}_{:,d} | \mathbf{X}, \Phi), \quad (3)$$

$$= \frac{1}{(2\pi)^{\frac{DN}{2}} |\mathbf{K}|^{\frac{D}{2}}} \exp \left(-\frac{1}{2} \text{tr}((\mathbf{K})^{-1} \mathbf{Y} \mathbf{Y}^T) \right)$$

where \mathbf{K} is the $N \times N$ covariance matrix obtained from the covariance function $k_{\text{ARD}}(\mathbf{x}, \mathbf{x}')$ and observation noise β , $\mathbf{Y}_{:,d}$ represent columns of the observation samples. A prior can be placed on the latent variables \mathbf{X} and marginalization w.r.t \mathbf{X} leads to a full Bayesian treatment,

$$p(\mathbf{X}) = \prod_{n=1}^N \mathcal{N}(\mathbf{x}_n | \mathbf{0}, \mathbf{I}),$$

$$p(\mathbf{Y} | \Phi) = \int p(\mathbf{Y} | \mathbf{X}, \Phi) p(\mathbf{X}) d\mathbf{X} \quad (4)$$

However, the integral for marginalization becomes intractable as \mathbf{X} appears non-linearly in the inverse of the kernel covariance matrix \mathbf{K} as shown in Eqn. (2),(3). To make the marginalization tractable, approximate variational inference can be applied wherein a variational distribution $q(\mathbf{X})$ is used to approximate the true posterior distribution $p(\mathbf{X} | \mathbf{Y})$ given by,

$$q(\mathbf{X}) = \prod_{n=1}^N \mathcal{N}(x_n | \mu_n, S_n) \quad (5)$$

where $\{\mu_n, S_n\}_{n=1}^N$ are the variational parameters. A Jensen's lower bound on the log marginal likelihood $\log p(\mathbf{Y})$ can be derived as follows:

$$F(q) = \int q(\mathbf{X}) \log \frac{p(\mathbf{Y} | \mathbf{X}) p(\mathbf{X})}{q(\mathbf{X})} d\mathbf{X} \quad (6)$$

The hyper parameters Φ are dropped for notational simplicity. The lower bound still remains intractable as the latent variables appear non-linearly in the conditional likelihood term $p(\mathbf{Y} | \mathbf{X})$.

Titsias *et al.* [5] resolved this problem by introducing data augmentation which is commonly used in sparse GP regression. Data augmentation involves adding M extra observations $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$, $\mathbf{u}_m \in \mathbb{R}^D$ known as inducing variables. These are evaluated at a set of pseudo inputs $\hat{\mathbf{X}} \in \mathbb{R}^{M \times L}$ through the same GP Prior as the latent variables, \mathbf{X} . The joint probability density and the variational distribution under this augmentation are modified as follows,

$$p(\mathbf{Y}, \mathbf{U}, \mathbf{X}, \hat{\mathbf{X}}) = p(\mathbf{Y} | \mathbf{U}, \mathbf{X}) p(\mathbf{U} | \hat{\mathbf{X}}) p(\mathbf{X}),$$

$$q(\Theta) = p(\mathbf{Y} | \mathbf{U}, \mathbf{X}) q(\mathbf{U}) q(\mathbf{X}) \quad (7)$$

where $q(\mathbf{X})$ takes the form of Eqn. (5), $q(\mathbf{U})$ is a variational distribution on the inducing variables whose form needs to be optimized and $p(\mathbf{Y} | \mathbf{U}, \mathbf{X})$ is the GP likelihood constrained by the latent variables as well as the inducing variables. This augmented probability model leads to a tractable Jensen's lower bound $\hat{F}(q)$ through the removal of the non-linear factor $p(\mathbf{Y} | \mathbf{X})$ thereby making the approximation tractable. Detailed derivations of the model are further presented in [5].

The prediction of unseen test data \mathbf{y}^* is performed by evaluating $p(\mathbf{y}^* | \mathbf{Y})$,

$$p(\mathbf{y}^* | \mathbf{Y}) = \frac{\int p(\mathbf{y}^*, \mathbf{Y} | \mathbf{x}^*, \mathbf{X}) p(\mathbf{x}^*, \mathbf{X}) d\mathbf{X} d\mathbf{x}^*}{\int p(\mathbf{Y} | \mathbf{X}) p(\mathbf{X}) d\mathbf{X}} \quad (8)$$

The predictive distribution is given by the ratio of two marginal likelihoods, both of which can be approximated using the augmented probability model i.e. $\exp(\hat{F}(q, \mathbf{X}, \mathbf{x}^*))$, $\exp(\hat{F}(q, \mathbf{X}))$. Efficient computations to handle test data are further described in [5].

Damianou *et al.* [7] proposed an extension to BGPLVM for learning a shared latent space among multiple observation spaces called Manifold Relevance Determination (MRD). Here we present MRD formulation for two observation spaces as shown in Fig. 3b i.e. $\mathbf{Y} \in \mathbb{R}^{N \times D_Y}$, $\mathbf{Z} \in \mathbb{R}^{N \times D_Z}$ assumed to be generated from a single latent variable $\mathbf{X} \in \mathbb{R}^{N \times L}$ through the GP mappings $f^Y : X \rightarrow Y$, $f^Z : X \rightarrow Z$,

$$\begin{aligned} \mathbf{y}_n &= f^Y(\mathbf{x}_n) + \epsilon_n^Y, \quad \epsilon_n^Y \in \mathcal{N}(\mathbf{0}, \beta_Y^{-1} \mathbf{I}), \\ \mathbf{z}_n &= f^Z(\mathbf{x}_n) + \epsilon_n^Z, \quad \epsilon_n^Z \in \mathcal{N}(\mathbf{0}, \beta_Z^{-1} \mathbf{I}) \end{aligned} \quad (9)$$

where $\epsilon_n^Y, \epsilon_n^Z$ are the noise random variables parametrized by the inverse variance parameters β_Y, β_Z . The GP mappings for the Y observation space can be modeled using the ARD kernel,

$$k_Y(\mathbf{x}_i, \mathbf{x}_j) = \sigma_Y^2 \exp \left(-\frac{1}{2} \sum_{l=1}^L \alpha_k^Y (x_{i,l} - x_{j,l})^2 \right) \quad (10)$$

and similarly for the Z observation space. Learning the ARD weights $\{\alpha_l^Y, \alpha_l^Z\}$ results in not only inferring the latent space dimensionality but also partitioning of the latent space into shared (\mathbf{X}_S) and private spaces ($\mathbf{X}_Y, \mathbf{X}_Z$). This is done using a heuristically set threshold δ on the normalized ARD weights to determine a latent dimension's relevance to each observation space,

$$\begin{aligned} \mathbf{X}_S &= \{\mathbf{x}_l\}_{l=1}^L : \mathbf{x}_l \in \mathbf{X}, \alpha_Y > \delta, \alpha_Z > \delta, \\ \mathbf{X}_Y &= \{\mathbf{x}_l\}_{l=1}^L : \mathbf{x}_l \in \mathbf{X}, \alpha_Y > \delta, \alpha_Z < \delta, \\ \mathbf{X}_Z &= \{\mathbf{x}_l\}_{l=1}^L : \mathbf{x}_l \in \mathbf{X}, \alpha_Y < \delta, \alpha_Z > \delta \end{aligned} \quad (11)$$

The objective is to evaluate the shared latent variables as well as the GP mapping hyper parameters for each observation space $\Phi\{\mathbf{Y}, \mathbf{Z}\}$. The joint conditional likelihood is obtained by factorizing each observation space as follows,

$$p(\mathbf{Y}, \mathbf{Z} | \mathbf{X}, \Phi_Y, \Phi_Z) = \prod_{\Gamma=\{\mathbf{Y}, \mathbf{Z}\}} p(\Gamma | \mathbf{X}, \Phi_\Gamma) \quad (12)$$

Marginalization of the latent variables, similar to BGPLVM, is intractable due to its non-linear appearance in the kernel covariance matrix. Damianou *et al.* [7] proposed an approximate variational inference formulation that relies on the use of an augmented probability model similar to BGPLVM (Eqn. (7)),

$$q(\Theta) = q(\mathbf{X}) \prod_{\Gamma=\{\mathbf{Y}, \mathbf{Z}\}} q(\mathbf{U}^\Gamma) p(f^\Gamma | \mathbf{U}^\Gamma, \mathbf{X}), \quad (13)$$

where $\mathbf{U}^{\{\mathbf{Y}, \mathbf{Z}\}}$ are the inducing variables for each observation space similar to the BGPLVM formulation. The Bayesian formulation further enables test inference in the form of $p(\mathbf{z}^* | \mathbf{y}^*)$ i.e. inference of test sample in Z observation space (\mathbf{z}_*) given the Y observation space \mathbf{y}^* . This inference is done by first estimating the latent sample \mathbf{x}^* similar to test inference given in Eqn. (8) and using this estimate through the GP mapping f^Z .

The inference for a test sample follows a sequence as shown in Fig. 3c. Firstly, the latent state x_Y^*, x_S^* corresponding to test sample y^* . The shared latent state x_S^* is then used to find nearest neighbors among the latent points corresponding to the

training data and obtain the private dimensions information for Z i.e. x_Z^{NN} . Finally, the full latent state x_S^*, x_Z^{NN} is used to infer the test pose state i.e. z^* . In this sequence, the computationally expensive operation is inference of x_Y^*, x_S^* as it involves optimization of marginal likelihoods similar to the MRD model training. This makes real-time inference difficult and so we explored alternative strategies for the test inference presented in Section III-E.

The latent variable models presented in this section are a powerful class of models that can be used in a wide range of settings. The use of GP mappings leads to data-efficient learning of complex mappings. Approximate Bayesian inference along with ARD kernels avoids overfitting and enables automatic dimensionality reduction.

B. Cloth State modeling

Clothing articles inherently lie in a high dimensional configuration space and feature extraction becomes a challenging task as the clothing article could have large shape variations and occlusions. Poor feature extraction could also lead to model inaccuracies for motor-skills learning thereby restricting the learning rate for robotic applications to cloth handling. To address this problem, we propose the use of Bayesian non-parametric latent variable models described in Section III-A. This leads to task-specific feature extraction in a purely data driven manner.

In this study, we consider clothing assistance tasks for demonstrating our proposed method. A mannequin was used as the subject and the clothing task is to cloth the mannequin with a T-shirt that is initially lying on its hands. We are interested in real-time estimation of the relationship between the assisted subject and cloth using a low-cost depth sensor for the implementation of a practical and efficient robotic clothing assistance framework. However, this is a challenging task as there are significant changes in the cloth state during clothing tasks along with self-occlusions and occlusion by the mannequin. To address this problem, we propose the learning of an offline cloth state model using information from both the depth sensor and motion capture system.

The purpose of the cloth model is to learn a latent representation $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$ corresponding to an aligned data set of clothing article observations from the depth sensor $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T$ and motion capture system $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]^T$. The motivation behind this modeling approach as described in Section I is that the motion capture system can provide precise location information of markers placed on the cloth where as the depth sensor can provide a generalized shape description. By learning a shared latent structure, we are indirectly learning a mapping from the generic depth sensor information to the more detailed motion capture information, which can be used for constrained cloth state estimation in real-time using noisy depth sensor observations. The predictive performance of the learned latent structure further depends on several factors such as the representations used for the observation spaces and the inference technique used. In the following subsections we describe the approach used in handling these factors.

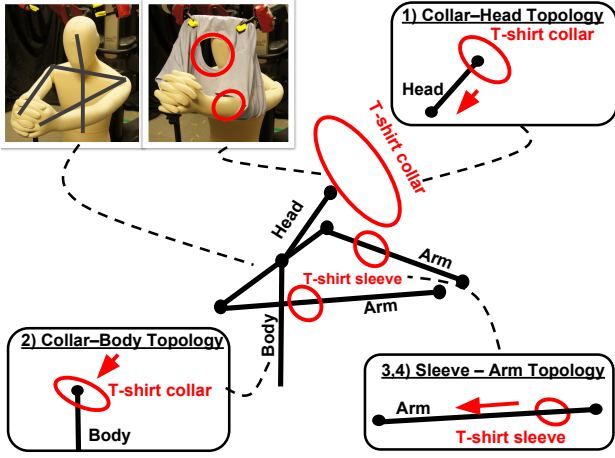


Fig. 4. Topology Coordinate representation used for the Human-Cloth Relationship as Pose Space Representation

C. Motion Capture Representations

The purpose of using motion capture system is to capture precise cloth state information which is required for efficient motor-skills learning. We consider the clothing task where the robot has to cloth a mannequin with a T-shirt which is initially on the mannequin's hands. We assume that the details of clothes such as wrinkles are not important to achieve clothing assistance tasks and hence used low-dimensional topology coordinates [8] to capture the relationship between human subject and clothing article. Furthermore, we have previously demonstrated that topology coordinates are robust to noise in the motion capture observations and can efficiently capture the human-cloth interaction in a practical setting [1].

Topology coordinates [8] were formulated for synthesizing human-like motions that involve close interactions. Topology coordinates compactly define the relationship between two curves in the Cartesian space using three different attributes, i.e. writhe w , center of twist $\mathbf{c} = [c_1 \ c_2]$ and density d . Writhe w measures the total twisting between two curves γ_1, γ_2 by using an approximation of the Gauss Linking Integral (GLI) [9]:

$$\text{GLI}(\gamma_1, \gamma_2) = \frac{1}{4\pi} \int_{\gamma_1} \int_{\gamma_2} \frac{(\gamma_1 - \gamma_2) \cdot (d\gamma_1 \times d\gamma_2)}{\|\gamma_1 - \gamma_2\|^3} \quad (14)$$

The center of twist \mathbf{c} , composed of two scalars explains the relative position of twist with respect to each of these lines. The density d represents the relative twisting between the two lines, i.e. which line is twisting around the other. These parameters can be analytically computed by dividing the given curves into chains of small line segments. The details for analytical computation of these parameters along with examples are presented in Appendix A.

The motor-skills required for the robot to complete the clothing task are 1) to pull the T-shirt collar over the mannequin's head and onto the mannequin's body, 2) to pull the T-shirt sleeves along the mannequin's arm towards its shoulder. To achieve these motor-skills, the following needs to be estimated and tracked by the depth sensor: T-shirt collar, T-shirt sleeves and the mannequin's posture. In this

study, the focus is on the cloth state estimation and therefore the mannequin's posture remains fixed during the task. The human-cloth relationship is given by considering the writhe and center of twist of the T-shirt w.r.t the mannequin for 4 different topologies as shown in Fig. 4: 1) T-shirt Collar - Mannequin's Head Topology, 2) Collar - Body, 3) Left Sleeve - Left Arm, 4) Right Sleeve - Right Arm thereby forming a 8 dimensional representation $\mathbf{Z} \in \mathbb{R}^8$. The density parameter is not considered as the T-shirt will always twist around the mannequin and this topology is never reversed for clothing assistance tasks.

The topology coordinate values were computed using the observations from motion capture system. The setup had eight infrared (IR) cameras placed carefully around the experimental setting to maximally avoid occlusion of markers. Six IR markers were attached on the T-shirt collar, three markers on each T-shirt sleeve and five markers on the mannequin respectively to estimate the human-cloth topological relationship. These markers were used to obtain approximate curves of the T-shirt collar and sleeves which were used in the computation of topology coordinates. The computation of topology coordinates for the T-shirt clothing task are presented in the Appendix A.

D. Depth Sensor Representations

A depth sensor is capable of capturing shape information of clothing articles. The purpose of the feature space representation is to capture the global cloth shape. For this, we consider the point cloud representation of the clothing article. The point cloud data can be used in real-time along with the proposed cloth modeling approach to infer precise human-cloth relationship information. In this section we present the method used to preprocess the RGB-D data and obtain the point cloud corresponding to the clothing article.

For the real-time estimation of human-cloth relationship, we need to track the overall cloth shape during the clothing task. There have been several studies that can reliably locate clothing articles within a cluttered environment [45], [46]. We assume that it is possible to obtain a seed bounding box for clothing articles through the existing methods. In this study, to simplify the process, we have used clothing articles that are of a single color to reliably localize the clothing article in an input frame.

The depth sensor provides a pair of RGB and depth images as each observation. The RGB image is used to locate the clothing article and the depth image is used to construct the cloth point cloud. Prior to the tracking, we perform a hue-saturation based color calibration where in a histogram of hue and saturation values is constructed from a region of interest (ROI) that corresponds to the T-shirt. This histogram can be used to find pixels corresponding to the T-shirt in an input image. For tracking of clothing article, we use the following approach as illustrated in Fig. 5:

- T-shirt hue-saturation histogram is applied to the input frame to obtain a back projection image. The back projection image is computed using the T-shirt histogram where the intensity of each pixel in the back-projection

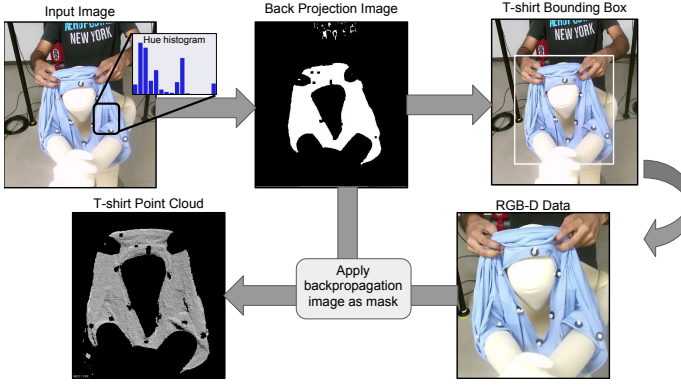


Fig. 5. Overview of the algorithm used to extract the point cloud corresponding to the T-shirt from raw RGB-D data from depth sensor.

image corresponds to the probability of belonging to the T-shirt.

- Back-projection image along with a seed T-shirt bounding box is provided as input to the standard CAMshift algorithm [10] where the shift and scaling of the bounding box between frames is estimated.
- To ensure feature consistency across multiple demonstrations, a bounding box of fixed size (250×250 in this case) is computed with a center corresponding to the bounding box obtained from the CAMshift algorithm.
- Back projection image within the bounding box represents the probability of belonging to the T-shirt and is applied as a mask to both the RGB and depth images and obtain the region corresponding to the T-shirt.
- Point cloud from the T-shirt depth image pixels is constructed using the intrinsic parameters of the sensor. This point cloud is further processed by applying statistical outliers removal techniques.

The image processing functions were implemented using the OpenCV library [11] and the point cloud processing was done using the Point Cloud Library (PCL) [12]. The point cloud constructed through preprocessing is down sampled to 50×50 forming a 7500 dimensional space with a triplet of 3 dimensions capturing the Cartesian position of a point in the T-shirt point cloud. $Y_{\text{depth}} \in \mathbb{R}^{7500}$

E. Real-time Implementation

The latent manifold learned by the MRD model includes two sets of ARD kernel weight parameters. These parameters describe the relevance of each latent dimension with respect to the corresponding observation space as described in Section III-A. The latent space is partitioned into three subspaces $\mathbf{X}_S, \mathbf{X}_Y, \mathbf{X}_Z$ where \mathbf{X}_S are the shared latent dimensions and $\mathbf{X}_Y, \mathbf{X}_Z$ are the private latent dimensions. The partitioning is done by placing manually set thresholds on the ARD weights as shown in Eqn. 11.

The objective of trained cloth models is to infer accurate pose state (motion capture space) given an unseen feature state (depth sensor space) in real-time. The inference of pose state z^* for an unseen feature state y^* involves a sequence of several steps as presented in Section III-A. As this inference approach

is not suitable for real-time implementation, we considered two alternate strategies with improved computational efficiency.

Optimization Approach: This is the standard strategy where optimization is performed for each test sample y^* to obtain the test latent state x_Y^*, x_S^* . This is the most computationally expensive approach and is expected to have the best predictive performance. *Nearest Neighbor Regression:* This is a naive strategy where we obtain the nearest neighbors y^{NN} to the test data y^* in the training set and approximate x_Y^*, x_S^* with mean of nearest neighbor latent points x_Y^{NN}, x_S^{NN} . This is the most computationally efficient approach and is expected to have the least predictive performance.

Hybrid Approach: This strategy can be considered as a trade-off between the two strategies presented above. The latent states obtained using the optimization approach were found to have strong temporal correlation. This insight was used to propose a hybrid inference strategy where an Unscented Kalman Filter (UKF) [14] is applied to latent states predicted using the nearest neighbor strategy. Furthermore, for every fixed number of observations, the internal state of UKF is updated using the optimization inference technique. This approach provides more reliable estimates compared to the nearest neighbor approach with similar computational efficiency and can be considered as a trade-off between accuracy and time complexity.

IV. EVALUATION

In this section, we describe the experiments conducted to evaluate the performance of our proposed framework. Section IV-A describes the experimental setup used and the dataset collected for evaluation. Section IV-B demonstrates the effectivity of using Bayesian nonparametrics in handling the non-rigidity of clothing articles. Section IV-C shows the predictive performance of the trained cloth models for various environmental settings. The computational complexity for the algorithm along with real-time implementation for test inference is demonstrated in Section IV-D.

A. Experimental Setup

Experimental setup includes the clothing assistance framework with Kinect V2 depth sensor and MAC3D motion capture system for cloth state estimation. We designed a framework using Robot Operating System (ROS) [15] and socket programming to integrate both devices and for synchronous data recording. Each sensor device has a program or node running on the control PC to perform data collection. We collected clothing trials with simultaneous observation of the T-shirt state using both the depth sensor and motion capture system. The node for motion capture system is designated as the master and other node as slave. Synchronization for data collection was performed by the master node which sends messages to the slave nodes for starting and stopping data collection. The observations were also temporally aligned ensuring point-to-point correspondences in the training phase. Cloth state is observed using both the sensors at a rate of 30 frames per second (FPS) during the clothing assistance

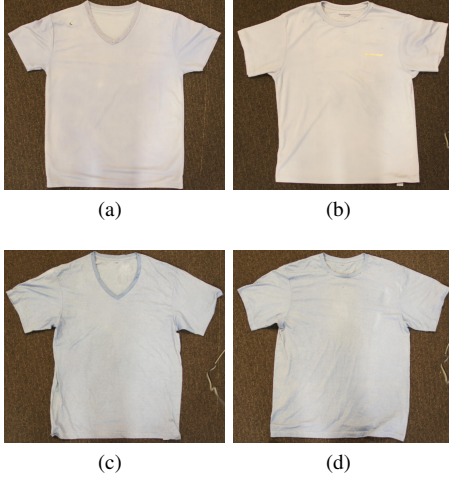


Fig. 6. Four T-shirts were used for collecting clothing trials: a) T-shirt 1: Polyester and V-neck, b) T-shirt 2: Polyester and round neck, c) T-shirt 3: Cotton and V-neck, d) T-shirt 4: Cotton and round neck

tasks. The observations were also spatially aligned by performing an absolute orientation calibration between the motion capture system and the depth sensor. The method proposed by Umeyama [17] was used to compute the transformation between the two reference frames. The source code for this framework is published online as a ROS package for further reference [16].

The efficiency of MRD to learn cloth state models was evaluated for clothing assistance tasks. Ideally, the learned cloth state model needs to be task specific such as for clothing tasks and should generalize to various environmental settings. For the case of robotic clothing assistance, the model needs to generalize to unseen postures of mannequin and different clothing materials. To evaluate the generalization capability, we used four T-shirts with different features as shown in Fig. 6. For each T-shirt, we collected clothing trials for six different postures of the mannequin obtained by varying the head inclination ($\{30^\circ, 45^\circ\}$) and the shoulder elevation ($\{100^\circ, 105^\circ, 110^\circ\}$) angles. The head inclination and shoulder elevation angles were measured with respect to the positive and negative Z-axis normal to the ground plane as shown in Fig. 7. A clothing demonstration along with the extracted feature representations and test inference for MRD is shown in the video demonstration (*framework.mp4*) included in the supplementary material.

The clothing trials in the dataset were collected through human demonstrations to ensure subtle variations which can not be induced by a robot in the demonstrations so that the dataset includes observations for different shapes of clothing articles. The motivation behind creating such a dataset was that the force applied by the robot changes largely for different T-shirts and different postures thereby imparting significant variations in the observed cloth state transitions across the clothing trials. The performance of using BGPLVM and MRD were evaluated using three metrics in all the experiments i.e. the Pearson correlation coefficient, root mean square error (RMSE) and normalized root mean square error (NRMSE).

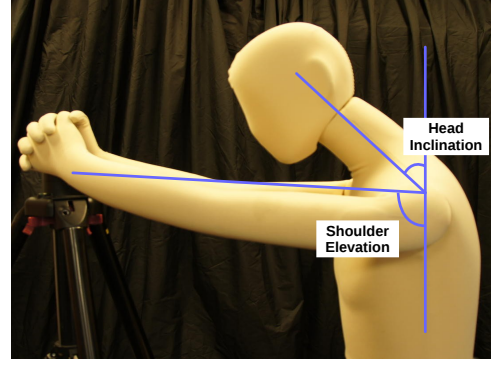


Fig. 7. Clothing trials were collected by varying the head inclination and shoulder elevation angles of the mannequin.

Given two univariate random variables x, y with samples $x_n, y_n : n \in \{1, \dots, N\}$ having means \bar{x}, \bar{y} , the metrics can be evaluated as follows:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{n=1}^N (x_n - y_n)^2}{N}} \\ \text{NRMSE} &= \frac{\text{RMSE}}{\max(\{x_n\}) - \min(\{x_n\})} \\ \text{Corr} &= \frac{\sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (x_n - \bar{x})^2} \sqrt{\sum_{n=1}^N (y_n - \bar{y})^2}} \end{aligned} \quad (15)$$

Statistical significance was evaluated for all the experiments by using the one-sided Wilcoxon signed rank sum test [23].

The training of an MRD model is a computationally expensive task which scales with the size of the training dataset as it is a kernel-based method. The computational complexity of the model scales as $O(NM^2)$ where N is the size of the training dataset and M are the number of inducing points used for the variational approximation. For all our experiments we have set the number of inducing points $M = 100$. We have conducted our experiments on a desktop machine with an Intel i7 3.5 GHz processor. The training time for an MRD model with 1275 observations took 3 hrs and 25 minutes for the model to converge. The BGPLVM and MRD models were trained using the GPy python library [20] along with the implementations for real-time inference. The source code to generate all the presented results are published online for further reference [21].

B. Latent Features Learned

In this section, we investigate the effectivity of using Bayesian nonparametrics and non-linear modeling for cloth state estimation. We evaluated the effectivity by only considering the depth sensor observation and by comparing the performance of BGPLVM with a linear latent variable model, Principal Component Analysis (PCA). We performed dimensionality reduction using both BGPLVM and PCA on the point cloud observation space and inspected the learned latent structures for both models. The BGPLVM model was optimized until there was negligible increments in the likelihood function and the variational distribution for the latent space was initialized

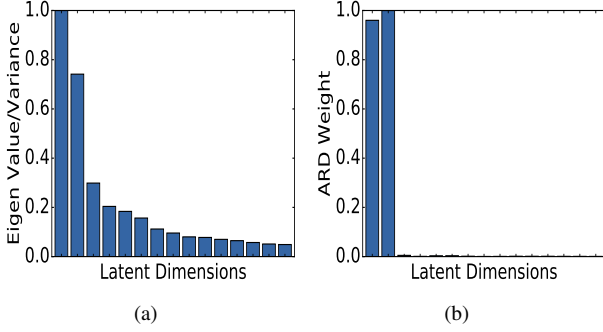


Fig. 8. Comparison of Latent Dimension relevance learned by PCA and BGPLVM: a) PCA relevance given by Eigen values, b) BGPLVM relevance given by ARD kernel weights

using the positions of training data in the latent space obtained from PCA.

For the observed data, we considered the point cloud representation from depth sensor as presented in Section III-D which is a 7500 dimensional observation space. The training data for the models was obtained from 5 clothing trials performed over 5 different postures with T-shirt 1 from Fig. 6. The test data for the model was given by 1 clothing trial for an unseen posture with T-shirt 1 and 3 clothing trials each using T-shirt 2,3 forming a total of 7 test clothing trials. Each clothing trajectory has about 100 samples measured at a frequency of 8 FPS, leading to 638 observations in the training data and 803 observations in the test data.

Fig. 8 demonstrate the relevance of each dimension in the latent space after training. The relevance for PCA is given by Eigen values and by the ARD kernel weight parameters for BGPLVM (Eqn. 2). The relevance parameters for both models are normalized such that the most significant dimension has a weight of 1.0 to demonstrate the relative importance between the dimensions. The relevance weights indicate that PCA takes all 15 dimensions to capture cloth transitions through the linear mapping, where as BGPLVM captures the underlying features within 2 dimensions using the non-linear GP mapping.

Fig. 9 shows the latent spaces for two most significant dimensions. The latent space learned by BGPLVM is constrained to a task-specific manifold in comparison to PCA. For BGPLVM, the samples from each clothing trial seem to follow a two dimensional latent trajectory which is consistent across the clothing trials with various environmental settings. The latent features explained by each dimension were inspected by reconstructing the high dimensional data for latent point variations only along that corresponding dimension. These dimensions explain the horizontal motion of the T-shirt collar and sleeves along the mannequin’s hands and the vertical motion on the T-shirt collar onto the mannequin’s head. An inspection of the latent space is included as a video demonstration (*bgplvm.mp4*) in the supplementary material wherein BGPLVM was applied to motion capture marker data and the latent feature represented by each dimension is evaluated by generating the high-dimensional marker space for changes along a single latent dimension.

To evaluate generalization capability, we compared recon-

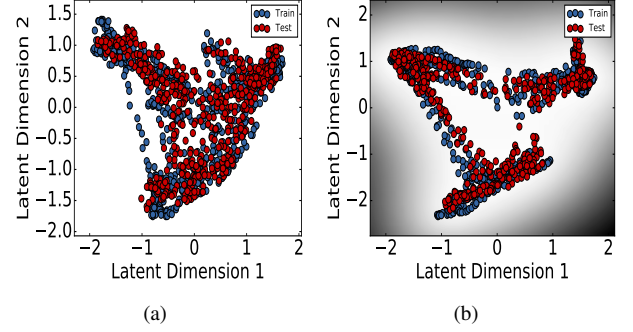


Fig. 9. Comparison of latent spaces learned by a) PCA and b) BGPLVM for the 1st, 2nd significant latent dimensions. Blue dots indicate training data, red dots indicate test data. Gray scale gradient for BGPLVM indicates the predictive variance obtained from the GP mapping.

struction error between PCA and BGPLVM as shown in Table I. Reconstruction error is given by comparing the input data point and the reconstructed data point from the latent point corresponding to the input,

$$\text{Err} = \|y_{\text{org}} - y_{\text{pred}}\|, \quad (16)$$

$$y_{\text{pred}} = f_{\text{model}}(f_{\text{model}}^{-1}(y_{\text{org}}))$$

where y_{org} is an input sample from dataset, y_{pred} is the predicted value after reconstruction and f_{model} is the forward mapping from latent space to observation space. We evaluated Root Mean Square Error (RMSE) and Pearson correlation as the metrics. The Wilcoxon signed rank sum test [23] was used to evaluate statistical significance and the p-value was evaluated for a one-sided test. We used an exact distribution over the W-statistic as the number of clothing trajectories were few (5 for training and 7 for test). For the training data, BGPLVM has much better performance as it is a kernel method and stores the complete training data unlike PCA. However, BGPLVM also has significantly better performance (p-value: 0.01) for the test data demonstrating its superior generalization capability for high dimensional and noisy point cloud data.

TABLE I
RECONSTRUCTION ERROR FOR PCA AND BGPLVM MODELS

Data	RMSE			Correlation		
	PCA	BGPLVM	p-Val	PCA	BGPLVM	p-Val
Train	0.016	0.012	0.05	0.646	0.800	0.05
Test	0.024	0.023	0.01	0.559	0.593	0.01

C. Predictive Performance of Cloth Models

Reliable cloth state estimation is a challenging problem due to the inherent ambiguity when observing from a single view point along with occlusion. We propose the learning of a shared latent manifold using Bayesian nonparametrics to disambiguate and solve the problem. In this section, we demonstrate the effectivity of using MRD for modeling cloth state. An MRD model is trained over observations from a

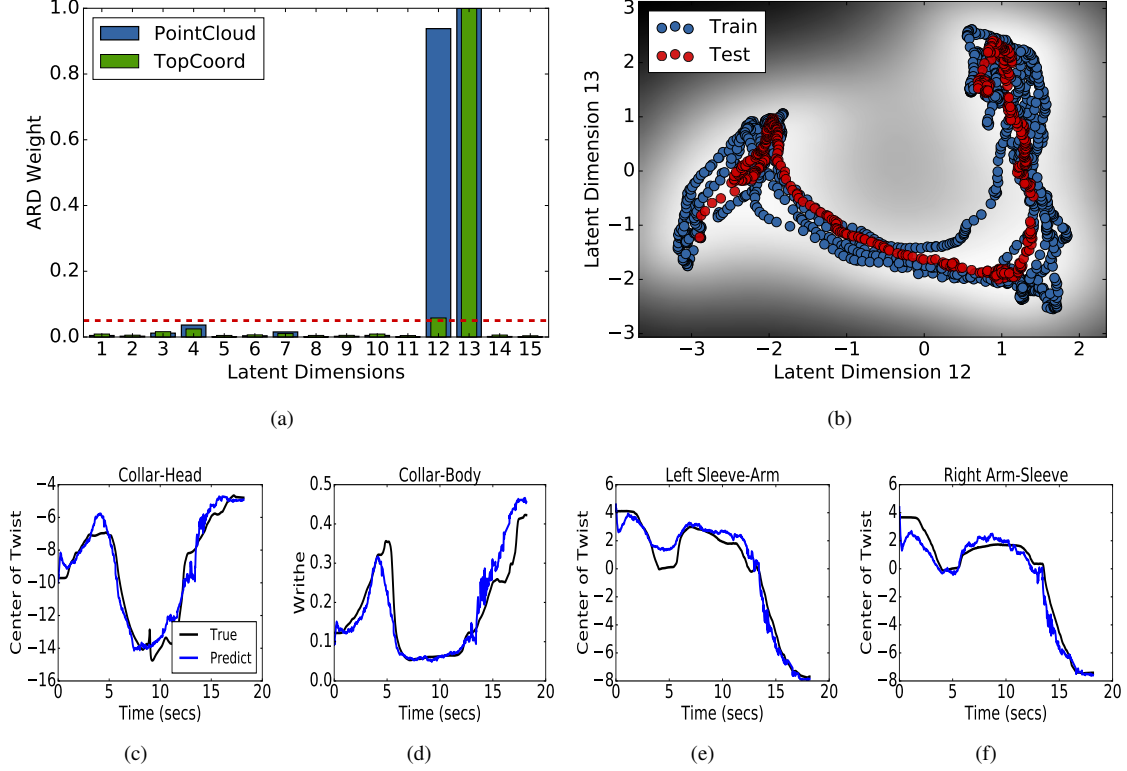


Fig. 10. MRD model trained for Point Cloud (\mathbb{R}^{7500}) and Topology Coordinate (\mathbb{R}^8) representations: a) ARD Kernel weights learned for each observation space, b) Latent manifold for two most significant dimensions with blue dots indicating training data and red indicating test data, c)-f) Inference of topology coordinates for unseen clothing trial.

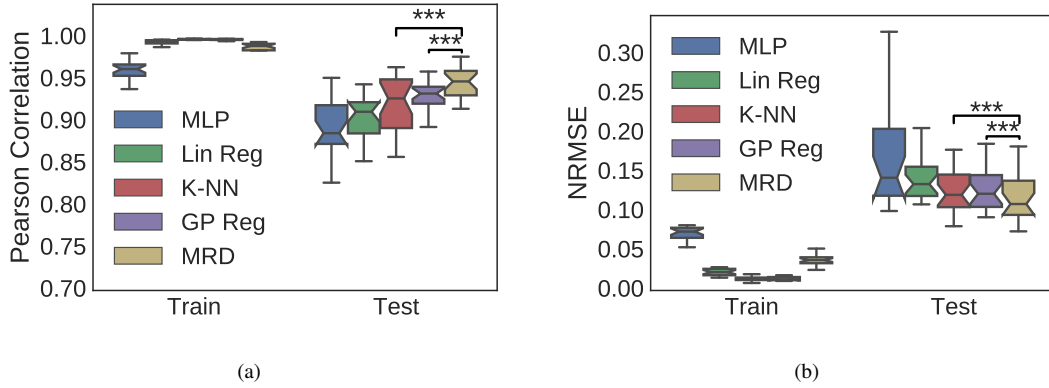


Fig. 11. Comparison of predictive performance between MRD and standard regression techniques. Evaluation on two metrics, a) Pearson correlation and b) Normalized RMS error. *** indicates $p < 0.001$ for one-sided Wilcoxon signed rank sum test [23].

depth sensor (feature space) and motion capture system (pose space). The trained model is then used to infer the human-cloth relationship information given a test cloth point cloud. Firstly, we present the latent features learned and the predictive performance of an example cloth model. We further present comparison of MRD with standard regression techniques to investigate the advantage of using a shared latent manifold for inferring cloth state.

Fig. 10 illustrates an MRD model between the topology coordinate and point cloud representations trained over 5 clothing trials for 5 different postures with T-shirt 1 of Fig. 6.

A clothing trial on an unseen posture is used as the test data. Fig. 10a shows the sets of ARD kernel weights that are learned. The threshold on ARD weights was set to 0.05 as shown by the red line leading to two shared dimensions between the observation spaces and no private dimensions for either observation space. However, this structure of the latent space, especially the private space dimensionality, was found to vary depending on the training data used. Fig. 10b shows the latent manifold for the two most significant dimensions. It can be seen that the model captures the dynamics of performing clothing tasks through the well-formed trajectories in the

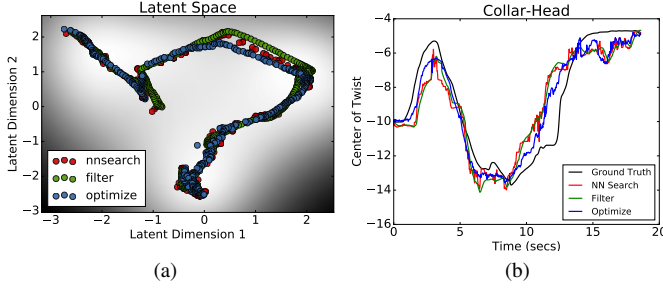


Fig. 12. a) Estimates for shared latent dimensions of MRD model evaluated for three inference strategies. b) Topology coordinate estimates from MRD model evaluated for three different inference strategies compared with the ground truth values

latent space. Fig. 10c-10f show the prediction of topology coordinate values for the test data where each figure indicates a particular topology presented in Section III-C demonstrating the predictive performance of MRD.

To validate the effectivity of using a shared latent manifold, we further compared the predictive performance of MRD with standard regression techniques. We considered four regression candidates i.e. linear regression, K nearest neighbor regression, multi-layer perceptron and Gaussian process regression. For nearest neighbor regression, we used 5 nearest neighbors for predictions. For neural networks, we used a single hidden layer with 200 hidden nodes and Rectified Linear Unit (ReLU) activation function in the network. GP regression was performed using the Radial Basis Function (RBF) kernel. All the models were trained until there were insignificant changes in the objective function for optimization.

The models were evaluated over a dataset of 24 clothing trials collected for 6 different postures of the mannequin using 4 T-shirts as described in Section IV-A. A set of 6 cloth models were trained for each T-shirt (total of 24 models) using leave-one-out cross validation wherein one clothing trial was used as test data and remaining 5 were used as training data. Fig. 11 shows the comparison of MRD with regression techniques. Normalized RMS error (Fig. 11b) and Pearson correlation (Fig. 11a) were used as the metrics for evaluation. Statistical significance was evaluated using the one-sided Wilcoxon signed rank sum test [23]. An approximate distribution over the W-statistic was used as the number of trials was relatively large ($N = 24$). It can be seen that MRD has the best predictive performance being significantly better ($p < 0.001$) over other regression techniques.

D. Comparison of Inference Methods

The inference for test data is a computationally expensive task that involves the optimization of a ratio of two marginal likelihoods similar to the training of the MRD model. To ensure real-time estimation of the human-cloth relationship, we considered two alternative strategies as presented in Section III-E. In this section, we present the relative predictive performance and computational complexity for these strategies. Our experimental setup was implemented such that, we could obtain raw T-shirt point cloud from the depth sensor and

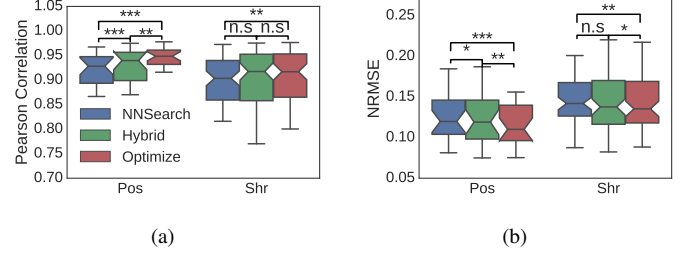


Fig. 13. Comparison of different inference strategies with MRD. Evaluation on two metrics, a) Pearson correlation and b) Normalized RMS error. *n.s.*: not significant, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$ for one-sided Wilcoxon signed rank sum test [23].

broadcast at a rate of 30 frames per second (fps) using the ROS framework. A separate program subscribes to the point cloud streams and infers the human-cloth relationship using one of the presented inference strategies.

The implementation details for each inference strategy is as follows. For the nearest neighbor search, the number of neighbors was set to 5 and a KD-tree [22] was used to search through the high dimensional training dataset. The optimization strategy was implemented as described in Section III-A with the initialization for latent point given by the nearest neighbor search. The hybrid strategy was applied by using an unscented Kalman filter to handle the non-linear transitions in the latent points obtained using the nearest neighbor strategy. The filter was only applied to the shared latent dimensions which vary from 2 to 4 depending on the latent manifold learned for different datasets. The state transition function was given by a constant velocity model with only the position used as observation variables. The internal state of UKF was updated every 15 observations which ensures a considerably good computational complexity. The parameters for the filter such as the process and measurement noise covariances were manually tuned minimizing the predictive error of the model.

We considered the performance of inference strategies for two different scenarios i.e. an unseen mannequin posture and for an unseen T-shirt. The MRD models were trained between the point cloud (feature space) and topology coordinate (pose space) representations. The evaluation dataset had 24 clothing trials as described in Section IV-C. For the unseen posture scenario, we performed leave one out cross-validation for each T-shirt with one posture as test data and the remaining five postures used for training. For the unseen T-shirt scenario, we used six clothing trials from 3 T-shirts as the training data and clothing trials for the unseen T-shirt as the test data. The state estimated for the shared latent dimensions along with the inferred topology coordinate values for an unseen T-shirt clothing trial is presented for the three inference strategies in Fig. 12a, 12b.

The performance of the inference strategies was evaluated using three metrics, i.e. normalized RMS error, Pearson correlation and the computational complexity as presented in Fig. 13. The nearest neighbor strategy has an average time complexity of processing 30 frames per second (fps), the hybrid method with a complexity of 10 fps and the

optimization method with about 1 fps. The results averaged over the test clothing trials have an intuitive trend, with nearest neighbor search having the least predictive performance and best computational complexity and the optimization based approach having the opposite trends. The hybrid approach is a good trade-off as it has a computational complexity suitable for a practical setting but with considerable improvements in the predictive performance. However, for the unseen T-shirt setting, the problem becomes quite difficult and there is no longer a significant improvement for the hybrid approach. This indicates the requirement for stronger temporal constraints such as placing a dynamics prior on the latent space as presented by Damianou *et al.* [6].

V. CONCLUSION

Assistance with clothing can greatly improve the quality of life as well as independence of the elderly and disabled. However, robotic assistance is still considered an open problem with several challenges involved. One of the challenge is the reliable estimation of human-cloth relationship which is crucial to ensure efficient learning of motor-skills and for the robot to reliably performing clothing tasks. In this study, we proposed the use of MRD to learn task-specific cloth state models that can be used for informed cloth state estimation. We implemented our framework to perform real-time estimation of human-cloth relationship using a low-cost depth sensor making it suitable for a feasible social implementation of robotic clothing assistance.

Clothing articles are non-rigid and inherently lie in a high-dimensional configuration space. We hypothesize that clothing articles undergo consistent deformations which vary from task to task and thereby lie in much lower task-specific latent space. We exploit these constraints and used Bayesian nonparametric latent variable models to learn the underlying latent features from high dimensional observation spaces. We used a low-cost depth sensor and the motion capture system to learn a shared latent manifold that captures complementary latent features from both systems in an offline manner and incorporated this shared model to reliably infer cloth state in real-time given high-dimensional and noisy depth sensor observations.

MRD is used to learn shared latent features between global cloth description given by depth sensor to precise human-cloth relationship parameters obtained using the motion capture system and infer the human-cloth relationship in real-time using this task-specific cloth model. We presented experimental results that demonstrate that BGPLVM is capable of learning consistent and meaningful latent features given noisy and high-dimensional observations of clothing articles. We further demonstrated the predictive performance and generalization capability of MRD for estimating the human-cloth relationship from noisy depth sensor readings. The effect of various factors such as observation space representations and inferences techniques was also shown.

The advantage of using MRD is that a corresponding latent space manifold can be learned for any observation space of the same underlying phenomenon. Based on this flexibility, our future work will be to learn models that incorporates T-shirt state as well as the assisted subject's posture and mainly

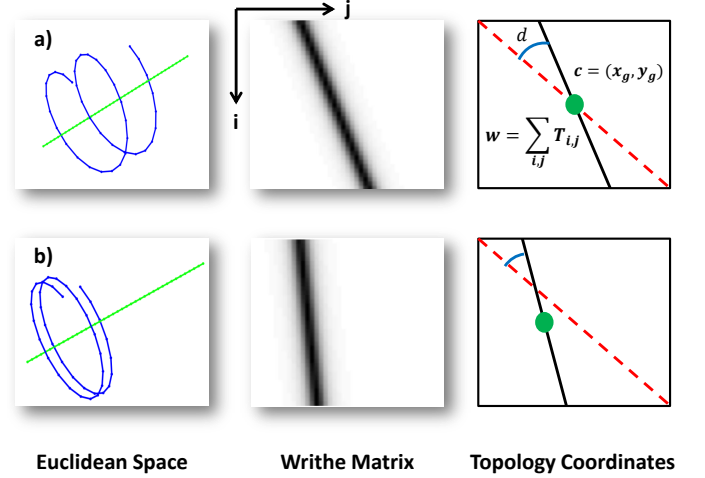


Fig. 14. Example showing the computation of topology coordinates for two configurations of Cartesian curves shown in Fig. 14a, 14b: (Left) Curves in Euclidean Space (Center) Writhe Matrix computed for both cases (Right) Topology coordinates computed from writhe matrix.

the proprioceptive information of the robot. This is based on the insight that while humans are performing clothing tasks they rely more on the forces experienced from the clothing article rather than the visual appearance of clothing articles. The reliable estimation of human-cloth relationship also enables the design of reinforcement-learning based motor-skills learning frameworks to ensure reliable performance of clothing tasks, where the robot can adapt to various types of failures faced during clothing trials.

APPENDIX A COMPUTATION OF TOPOLOGY COORDINATES

This appendix summarizes the computation of topology coordinates, as presented in [8]. Given two continuous curves γ_1 and γ_2 in the euclidean space, the topology coordinates are computed by approximating the Gaussian Linking Integral (GLI). Firstly the curves need to be divided into a number of small line-segments. These segments are used for constructing a writhe matrix $T \in \mathbb{R}^{N_1 \times N_2}$ where N_1, N_2 are the number of line segments in the curves.

Let r_{ab}, r_{cd} be two segments (one from each curve), where $a, b, c, d \in \mathbb{R}^3$ are the end points of the segments. Firstly the following vectors are calculated:

$$\begin{aligned} n_a &= \frac{r_{ac} \times r_{ad}}{\|r_{ac} \times r_{ad}\|}, & n_b &= \frac{r_{ad} \times r_{bd}}{\|r_{ad} \times r_{bd}\|} \\ n_c &= \frac{r_{bd} \times r_{bc}}{\|r_{bd} \times r_{bc}\|}, & n_d &= \frac{r_{bc} \times r_{ac}}{\|r_{bc} \times r_{ac}\|} \end{aligned} \quad (17)$$

Using these vectors, the writhe between the line segments is given by:

$$T_{i,j} = \arcsin(n_a^T n_b) + \arcsin(n_b^T n_c) + \arcsin(n_c^T n_d) + \arcsin(n_d^T n_a) \quad (18)$$

where $T_{i,j}$ is the (i, j) th element in the writhe matrix. Now the

- Writhe w of the two curves is given by summation over the writhe matrix as a measure of the total amount of

twisting between the curves: topology coordinates are computed from the writhe matrix as follows:

$$w = \text{GLI}(\gamma_1, \gamma_2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} T_{i,j} \quad (19)$$

where N_1, N_2 are the number of segments for curves γ_1 and γ_2 respectively.

- The center of twist \mathbf{c} is given by two scalar values which indicate the center of twist for each curve about the other curve. These values are given by weighted summations of the writhe matrix:

$$\begin{aligned} \mathbf{c} &= (x_g, y_g) \\ &= \left(\frac{\sum_i^{N_1} \sum_j^{N_2} i T_{i,j}}{w} - \frac{N_2}{2}, \frac{\sum_i^{N_1} \sum_j^{N_2} j T_{i,j}}{w} - \frac{N_1}{2} \right) \end{aligned} \quad (20)$$

- The density d is given by computing the angle between the principal axis of the writhe matrix and the diagonal line of the matrix.

An example of computing the topology coordinates is shown in Fig. 14. The pane on the left shows two examples of curves in the euclidean space, the pane in the center shows the writhe matrix computed for each case and the pane on the right shows the computation of topology coordinates from the writhe matrix thereby forming the topology space. It can be seen from this example that the parameters change based on the topological relationship between the curves thereby capturing the complex relationship using few scalar parameters.

For clothing assistance tasks, the motion capture system was used to obtain the Cartesian position of markers placed on the mannequin and T-shirt. The markers placed on the mannequin were used to approximate its posture using a stick figure. For example, the left arm of the mannequin was approximated by a line segment joining two markers placed on its wrist and shoulder joint. Each line segment for the mannequin were divided into 20 segments $N_{\text{mannequin}} = 20$. The markers placed on the T-shirt were used to obtain circle approximations of its collar and sleeve shapes. The T-shirt collar curve was approximated using 40 segments and each of the sleeves were approximated with 20 segments i.e. $N_{\text{collar}} = 40, N_{\text{sleeves}} = 20$. This data was then used to compute the topology coordinates given by the equations presented above.

APPENDIX B

COMPARISON OF FEATURE REPRESENTATIONS

The latent features and predictive performance of MRD depends on the feature representations used for the observation spaces. In this appendix, we consider several representations for each observation space that are relevant to the clothing assistance framework and evaluate their relative performance. The appendix is divided into three sub sections. Firstly, we define alternate representations for the motion capture space and then for the depth sensor space. Finally we present experimental results to evaluate performance of these representations.

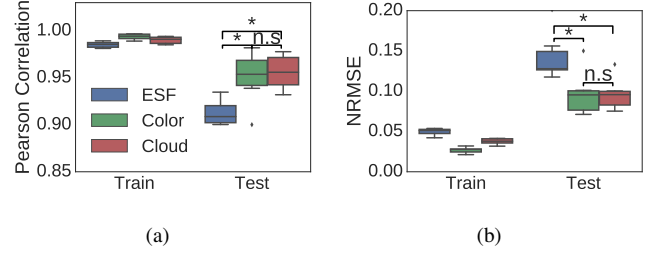


Fig. 15. Comparison of different feature representations for depth sensor observation space. Evaluation on two metrics, a) Pearson correlation and b) Normalized RMS error. *n.s.*: not significant, *: $p = 0.025$ for one-sided Wilcoxon signed rank sum test [23].

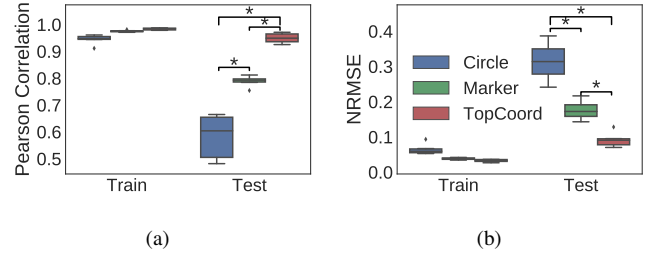


Fig. 16. Comparison of different feature representations for motion capture observation space. Evaluation on two metrics, a) Pearson correlation and b) Normalized RMS error. *n.s.*: not significant, *: $p = 0.025$ for one-sided Wilcoxon signed rank sum test [23].

A. Motion Capture Representations

To evaluate the effect of pose space representation on the predictive performance of the cloth model, we considered two alternative representations along with topology coordinates:

- **Marker Representation:** given by the Cartesian position of each of the 12 markers placed on the collar and sleeves of the T-shirt forming a 36 dimensional space.
- **Circle Approximation:** given by the parameters of a circle approximation to the T-shirt collar and the sleeves obtained from the marker positions. Each circle is parametrized by $[C \ r \ \vec{n}] \in \mathbb{R}^7$ i.e. its center ($C \in \mathbb{R}^3$), radius ($r \in \mathbb{R}$) and normal ($\vec{n} \in \mathbb{R}^3$) thereby forming a 21 dimensional feature representation.

B. Depth Sensor Representations

We also considered two alternate representations for the RGB-D data to evaluate the effect of feature space representation. Each representation for the feature space captures different physical aspect of shape information.

- **Color Pixel Data:** The color pixel data from the bounding box of the T-shirt is evenly down sampled to 50×50 and converted to single channel thereby forming a 2500 dimensional space with dimension being the color intensity of each pixel. $Y_{\text{color}} \in \mathbb{R}^{2500}$
- **Ensemble of Shape Functions:** ESF is a global feature descriptor proposed by Wohlkinger *et al.* [13] that is primarily used to represent the underlying shape of a 3D point cloud. ESF is a fixed 640 dimensional feature histogram, consisting of a concatenation of 10 histograms with 64 bins each in them. These histograms are generated by

repeated random sampling of pairs or triplets of points from the point cloud and computing various parameters of the resultant triangles and lines. $Y_{ESF} \in \mathbb{R}^{640}$

C. Evaluation

The feature representations used for an observation space capture different information about clothing articles thereby leading to different latent features on training. For this purpose, we used several representations for each observation space as presented in Section III-D, III-C. We compared the predictive performance for each of the representation to evaluate the best representation in each observation space for clothing assistance tasks. The representations for the feature space were trained along with topology coordinate representation for comparison and the pose space representations were trained along with the point cloud representation. Observations from 6 clothing trials for 6 different postures of the mannequin with T-shirt 1 were used as the evaluation dataset. We performed 6-fold cross-validation on the dataset with a single clothing trial taken as test data and the remaining 5 trials used as training data.

The results for comparison between the representations is shown in Fig. 15, 16 given by the mean values of normalized RMS error and Pearson correlation across the 6-folds. Fig. 15 indicates that the point cloud representation and the color representations have almost similar performance and are significantly better than the ESF representation. This indicates that both color or point cloud representations are suitable for the task of clothing assistance. However, the ESF representation being a feature histogram seems to drop some crucial shape information that is necessary for reliable cloth state estimation. Fig. 16 indicates that the topology coordinate representation has the best predictive performance. The marker and circle approximation representations have significantly lower performance ($p = 0.025$) as the variation between clothing trials for specific marker positions is significantly higher in comparison to topology coordinates which mainly captures the relationship between human and cloth rather than specific cloth state.

APPENDIX C GENERALIZABILITY OF CLOTH MODELS

In this appendix, we evaluate the generalizability of cloth models trained using MRD for various environmental settings. Ideally, we would want the cloth model to learn clothing task specific latent features and generalize to unseen postures of the mannequin and unseen clothing articles. To evaluate the generalization capability, we conducted two sets of experiments. In the first experiment, we evaluated the generalization to unseen postures. For this we considered 4 sets of 6 clothing trials from 4 T-shirts and 6 postures each. We performed 6-fold cross validation across postures for each T-shirt and evaluated the predictive performance as shown in Fig. 17. The results show that the trained cloth models can generalize well to unseen mannequin postures with mean normalized RMSE 0.134 and mean Pearson correlation 0.925.

In the second experiment, we evaluated generalization to unseen T-shirts. We considered a similar dataset as the posture

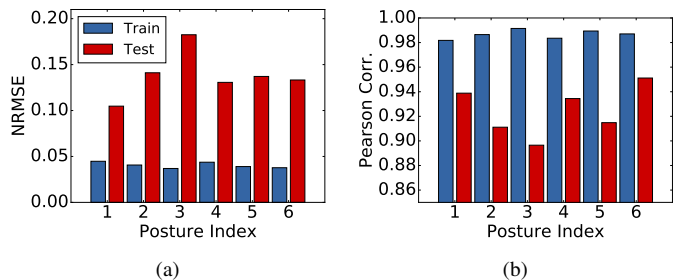


Fig. 17. Generalization capability of cloth model to unseen postures evaluated through 6-fold cross validation given by a) Normalized RMSE and b) Pearson Correlation values

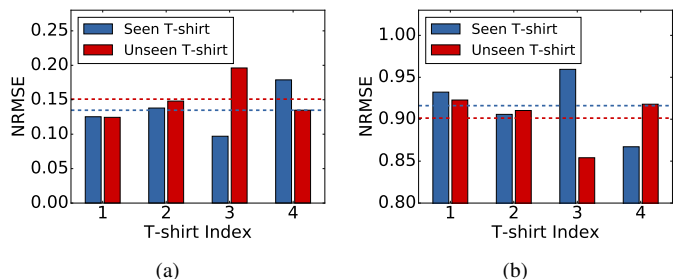


Fig. 18. Generalization to clothing trials for seen and unseen T-shirts given by a) Normalized RMSE and b) Pearson Correlation values. Horizontal dotted lines indicate the mean values across the T-shirts.

experiment, however we performed 4-fold cross validation across the T-shirts and evaluated the predictive performance to unseen T-shirts as shown in Fig. 18. For each cross-validation, we included 6 clothing trials from 3 T-shirts and trained an MRD model. The predictive performance was evaluated for unseen clothing trials of both seen and unseen T-shirts. The results indicate that the performance is slightly better for seen T-shirts however the performance is also good for unseen T-shirts.

ACKNOWLEDGMENT

This work was supported in part by the Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science (No. 16H01749). This work was also supported in part by IMPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan) 2015-PM07-36-01.

REFERENCES

- [1] Nishanth Koganti, Tomoya Tamei, Takamitsu Matsubara, and Takuma Shibata. Real-time estimation of human-cloth topological relationship using depth sensor for robotic clothing assistance. In *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*, pages 124–129. IEEE, 2014.
- [2] Nishanth Koganti, Jimson Gelbolingo Ngeo, Tamei Tomoya, Kazushi Ikeda, and Tomohiro Shibata. Cloth dynamics modeling in latent spaces and its application to robotic clothing assistance. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 3464–3469. IEEE, 2015.
- [3] Tomoya Tamei, Takamitsu Matsubara, Akshara Rai, and Tomohiro Shibata. Reinforcement learning of clothing assistance with a dual-arm robot. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 733–738. IEEE, 2011.

- [4] Christopher KI Williams and Carl Edward Rasmussen. Gaussian processes for machine learning. *the MIT Press*, 2006.
- [5] Michalis K Titsias and Neil D Lawrence. Bayesian gaussian process latent variable model. In *International Conference on Artificial Intelligence and Statistics*, pages 844–851, 2010.
- [6] Andreas Damianou, Michalis K Titsias, and Neil D Lawrence. Variational gaussian process dynamical systems. In *Advances in Neural Information Processing Systems*, pages 2510–2518, 2011.
- [7] Andreas Damianou, Carl Ek, Michalis Titsias, and Neil Lawrence. Manifold relevance determination. *arXiv preprint arXiv:1206.4610*, 2012.
- [8] Edmond SL Ho and Taku Komura. Character motion synthesis by topology coordinates. In *Computer Graphics Forum*, volume 28, pages 299–308. Wiley Online Library, 2009.
- [9] William F Pohl. The self-linking number of a closed space curve(gauss integral formula treated for disjoint closed space curves linking number). *Journal of Mathematics and Mechanics*, 17:975–985, 1968.
- [10] Gary R Bradski. Computer vision face tracking for use in a perceptual user interface. 1998.
- [11] Gary Bradski et al. The opencv library. *Doctor Dobbs Journal*, 25(11):120–126, 2000.
- [12] Aitor Aldama, Zoltan-Csaba Marton, Federico Tombari, Walter Wohlking, Christian Potthast, Bernhard Zeisl, Radu Bogdan Rusu, Suat Gedikli, and Markus Vincze. Point cloud library. *IEEE Robotics & Automation Magazine*, 1070(9932/12), 2012.
- [13] Walter Wohlking and Markus Vincze. Ensemble of shape functions for 3d object classification. In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, pages 2987–2992. IEEE, 2011.
- [14] Eric A Wan and Ronell Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. IEEE, 2000.
- [15] Morgan Leigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5, 2009.
- [16] Nishanth Koganti. Robotic clothing assistance interface. https://github.com/ShibataLabPrivate/cloth_assist_interface, 2015–2017.
- [17] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):376–380, 1991.
- [18] Wikipedia. Pearson product-moment correlation coefficient — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Pearson%20product-moment%20correlation%20coefficient&oldid=702729237>, 2016. [Online; accessed 03-February-2016].
- [19] Wikipedia. Root-mean-square deviation — Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Root-mean-square_deviation, 2016. [Online; accessed 03-February-2016].
- [20] The GPy authors. GPy: A gaussian process framework in python. <http://github.com/SheffieldML/GPy>, 2012–2015.
- [21] Nishanth Koganti. Bayesian nonparametric learning of cloth models. <https://github.com/buntkye/TRo2017>, 2016–2017.
- [22] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.
- [23] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.
- [24] Stephen Miller, Mario Fritz, Trevor Darrell, and Pieter Abbeel. Parametrized shape models for clothing. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4861–4868. IEEE, 2011.
- [25] Stephen Miller, Jur Van Den Berg, Mario Fritz, Trevor Darrell, Ken Goldberg, and Pieter Abbeel. A geometric approach to robotic laundry folding. *The International Journal of Robotics Research*, 31(2):249–267, 2012.
- [26] Arnau Ramisa, Guillem Alenya, Francesc Moreno-Noguer, and Carme Torras. Finddd: A fast 3d descriptor to characterize textiles for robot manipulation. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 824–830. IEEE, 2013.
- [27] Arnau Ramisa, Guillem Alenya, Francesc Moreno-Noguer, and Carme Torras. Learning rgb-d descriptors of garment parts for informed robot grasping. *Engineering Applications of Artificial Intelligence*, 35:246–258, 2014.
- [28] Bryan Willimon, Stan Birchfield, and Ian Walker. Classification of clothing using interactive perception. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1862–1868. IEEE, 2011.
- [29] Bryan Willimon, Stan Birchfield, and Ian Walker. Interactive perception of rigid and non-rigid objects. *Int J Adv Robotic Sy*, 9(227), 2012.
- [30] Yasuyo Kita, Fumio Kanehiro, Toshio Ueshiba, and Nobuyuki Kita. Clothes handling based on recognition by strategic observation. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 53–58. IEEE, 2011.
- [31] Yasuyo Kita, Toshio Ueshiba, Fumio Kanehiro, and Nobuyuki Kita. Recognizing clothing states using 3d data observed from multiple directions. In *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, pages 227–233. IEEE, 2013.
- [32] Alexandros Doumanoglou, Andreas Kargakos, Tae-Kyun Kim, and Sotiris Malassiotis. Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 987–993. IEEE, 2014.
- [33] Sandy H Huang, Jia Pan, George Mulcaire, and Pieter Abbeel. Leveraging appearance priors in non-rigid registration, with application to manipulation of deformable objects. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 878–885. IEEE, 2015.
- [34] Karthik Lakshmanan, Apoorva Sachdev, Ziang Xie, Dmitry Berenson, Ken Goldberg, and Pieter Abbeel. A constraint-aware motion planning algorithm for robotic folding of clothes. In *Experimental Robotics*, pages 547–562. Springer, 2013.
- [35] Steven D Klee, Beatriz Quintino Ferreira, Rui Silva, Joao Paulo Costeira, Francisco S Melo, and Manuela Veloso. Personalized assistance for dressing users. In *Social Robotics*, pages 359–369. Springer, 2015.
- [36] A. Colome, A. Planells, and C. Torras. A friction-model-based framework for reinforcement learning of robotic tasks in non-rigid environments. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 5649–5654, 2015.
- [37] Yixing Gao, Hyung Jin Chang, and Yiannis Demiris. User modelling for personalised dressing assistance by humanoid robots. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 1840–1845. IEEE, 2015.
- [38] Kinya Yamazaki, Ryosuke Oya, Kazuhiro Nagahama, and Masayuki Inaba. A method of state recognition of dressing clothes based on dynamic state matching. In *System Integration (SII), 2013 IEEE/SICE International Symposium on*, pages 406–411. IEEE, 2013.
- [39] Kimitoshi Yamazaki, Ryosuke Oya, Kotaro Nagahama, Kei Okada, and Masayuki Inaba. Bottom dressing by a life-sized humanoid robot provided failure detection and recovery functions. In *System Integration (SII), 2014 IEEE/SICE International Symposium on*, pages 564–570. IEEE, 2014.
- [40] Aaron Shon, Keith Grochow, Aaron Hertzmann, and Rajesh P Rao. Learning shared latent structure for image synthesis and robotic imitation. In *Advances in Neural Information Processing Systems*, pages 1233–1240, 2005.
- [41] Sebastian Bitzer, Matthew Howard, and Sethu Vijayakumar. Using dimensionality reduction to exploit constraints in reinforcement learning. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 3219–3225. IEEE, 2010.
- [42] Jonathan Ko and Dieter Fox. Learning gp-bayesfilters via gaussian process latent variable models. *Autonomous Robots*, 30(1):3–23, 2011.
- [43] Zhikun Wang, Katharina Mülling, Marc Peter Deisenroth, Heni Ben Amor, David Vogt, Bernhard Schölkopf, and Jan Peters. Probabilistic movement modeling for intention inference in human–robot interaction. *The International Journal of Robotics Research*, 32(7):841–858, 2013.
- [44] Kevin Sebastian Luck, Gerhard Neumann, Erik Berger, Jan Peters, and Heni Ben Amor. Latent space policy search for robotics. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1434–1440. IEEE, 2014.
- [45] Xiao Wu, Bo Zhao, Ling-Ling Liang, and Qiang Peng. Clothing extraction by coarse region localization and fine foreground/background estimation. In *International Conference on Multimedia Modeling*, pages 316–326. Springer, 2013.
- [46] Kimitoshi Yamazaki. A method of classifying crumpled clothing based on image features derived from clothing fabrics and wrinkles. *Autonomous Robots*, pages 1–15, 2016.



Nishanth Koganti received B.Tech and M.E. degrees from Indian Institute of Technology Jodhpur, India and Nara Institute of Science and Technology, Japan in 2012 and 2014 respectively. He is currently a Ph.D student in Nara Institute of Science and Technology, Japan. His research interests are assistive robotics, motor-skills learning, and machine learning. He received the IROS 2015 Best Application Paper Award (2015).



Tomoya Tamei received his Ph.D. degree from Nara Institute of Science and Technology in 2009. He is an assistant professor in the Graduate School of Information Science, Nara Institute of Science and Technology, Japan. His research interests are robotics, motor neurophysiology, and machine learning. He received the best paper award from the Japanese Neural Network Society (2015) and IROS 2015 Best Application Paper Award (2015).



Kazushi Ikeda received his B.E., M.E., and Ph.D. degrees in mathematical engineering and information physics from the University of Tokyo in 1989, 1991, and 1994, respectively. He was a research associate with the Department of Electrical and Computer Engineering, Kanazawa University from 1994 to 1998. In 1995, he was a research associate of the Chinese University of Hong Kong for three months. From 1998 to 2008, he was with the Graduate School of Informatics, Kyoto University, as an associate professor. Since 2008, he has been a full

professor of Nara Institute of Science and Technology. He is the editor-in-chief of Journal of Japanese Neural Network Society, an action editor of Neural Networks, an associate editor of IEEE Transactions on Neural Networks and Learning Systems.



Tomohiro Shibata received his B.E., M.E., and Ph.D. degrees in 1991, 1993, and 1996 from the University of Tokyo. He is currently a professor at the Graduate School of Life Science and Systems Engineering of the Kyushu Institute of Science and Technology. His main research interest is on understanding and assisting motor control and decision making by humans by using interdisciplinary approaches. He received a young investigator award from the Robotics Society of Japan (1992), the best paper award from the Japanese Neural Network

Society (2002, 2015), the Neuroscience Research Excellent Paper Award from the Japan Neuroscience Society (2007), the ICROS Award for IROS 2015 Best Application Paper (2015). He is a member of IEEE Robotics and Automation Society, IEEE Computational Intelligence Society, the Robotics Society of Japan, Robotics Society of India, the Institute of Electronics, Information and Communication Engineers, and Japanese Neural Network Society. He is also an Editorial Board Member of Neural Networks, and an executive board member of the NPO Agora Music Club.